

Second-law-like inequalities with information and their interpretations

Jordan M. Horowitz¹ and Henrik Sandberg²

¹ Department of Physics, University of Massachusetts at Boston, Boston, MA 02125, USA

² Department of Automatic Control, KTH Royal Institute of Technology, Stockholm, Sweden

E-mail: Jordan.Horowitz@umb.edu

Abstract. In a thermodynamic process with measurement and feedback, the second law of thermodynamics is no longer valid. In its place, various second-law-like inequalities have been advanced that each incorporate a distinct additional term accounting for the information gathered through measurement. We quantitatively compare a number of these information measures using an analytically tractable model for the feedback cooling of a Brownian particle. We find that the information measures form a hierarchy that reveals a web of interconnections. To untangle their relationships, we address the origins of the information, arguing that each information measure represents the minimum thermodynamic cost to acquire that information through a separate, distinct measurement protocol.

Keywords: nonequilibrium thermodynamics, feedback, information theory, optimal control theory

Submitted to: *New Journal of Physics*

1. Introduction

The Kelvin-Planck statement of the second law of thermodynamics forbids the existence of a cyclically operating device whose sole effect is to convert heat from a single thermal reservoir into an equal amount of work [1]. However, we can circumvent this restriction, if our device operates via measurement and feedback: a possibility first envisioned by Szilard in his famous thought experiment [2]. Recently, there has been renewed interest in this old idea spurred by the development of a collection of distinct, second-law-like inequalities that quantify the interplay between the information gathered through measurement and the work that can be extracted in response through feedback. For continuously operating devices at temperature T , all these predictions bound the extracted work rate \dot{W}_{ext} as

$$\dot{W}_{\text{ext}} \leq k_{\text{B}} T \dot{I}, \quad (1)$$

by some information acquisition rate, generically denoted here as \dot{I} , which differs in each second-law-like inequality, and k_{B} is Boltzmann's constant. The first inequality of this form was derived by Sagawa and Ueda for a single feedback loop [3], but subsequently has been extended to include the repeated use of feedback, allowing for the application to continuously operating information engines [4, 5, 6, 7, 8, 9, 10, 11]. In this case, the information rate is identified as the rate of growth of the transfer entropy [12] from the system to the measurement device (or feedback controller) [8, 10, 13, 14]. An alternative inequality identifies the information rate with the flow of mutual information between the system and a continuously-interacting auxiliary measurement device. This information flow approach has been developed for small systems modeled as continuous diffusion processes [15], discrete Markov jump processes [14, 16], and for stochastic processes interacting discretely [17, 18]. Yet another version has been suggested by Kim and Qian specifically for the feedback cooling of a harmonically-trapped Brownian particle, where the extracted work is bounded by a term they call entropy pumping [19]. To date there is no clear information-theoretic interpretation of this term. Nevertheless, this result conforms to the second-law-like structure in (1). Further developments in this direction are the inclusion of measurement errors and delay [20, 21, 22]. At first glance, this plethora of seeming similar predictions is confusing and raises questions about the interpretation as well as the utility of these information bounds. To help clarify the situation, a number of studies have compared some of these measures from different points of view [14, 15, 23, 24]. Our goal in this paper is to build on these works by providing a comprehensive, pedagogical comparison of all these information measures within a single framework in order to show clearly their relationships and limitations.

There are essentially two ways to view (1). The first is to treat (1) simply as a numerical bound on the extracted work \dot{W}_{ext} without reference to the physical underpinnings of \dot{I} . This is the point of view we typically take when investigating feedback (or information) engines [25, 26, 27, 28, 23, 29], where our goal is to optimally extract the maximum amount of work; the maximum being any or all of the possible information measures. In this respect, having so many bounds is problematic, since

we are unsure which is the most appropriate. Nevertheless, this is the approach we take in the first half of our paper in section 3. There we investigate the quantitative relationship between the various information measures by analytically calculating them in a Brownian particle model of feedback cooling, which we introduce in section 2. We use this particular model, since it has been studied theoretically [19, 20, 21] and could be implemented experimentally in the setups of [30, 31]. The analytical tractability of this model further lets us examine these information measures from the point of view of optimal control theory, which reveals intimate connections among them. The second way to interpret (1) is to take seriously its resemblance to the second law, and ask how far can we push this analogy? In particular, the traditional statement of the second law dictates that the entropy production of the universe – system *and* surroundings – during a thermodynamic process must be positive [1]. In feedback-driven systems, the surroundings not only include the traditional thermodynamic reservoirs, such as heat baths or chemical baths, but in addition they include an auxiliary system that records the measurement and feeds back that information. In this case, does (1) still represent the entropy production of the system and its surroundings, except now the surroundings contain the feedback device? This is the question we address in the second half of our paper in section 4. There we observe that the transfer entropy rate and information flow have clear interpretations as the minimum entropy production required to acquire that information. However, each one is associated with a different physical measurement scenario, that is with a distinct surroundings in much the same way a particle reservoir differs from a thermal reservoir.

2. Feedback cooling model

Throughout, we will illustrate the different information concepts with a model for the feedback cooling of an underdamped Brownian particle [19, 20, 21]. This will allow us to discuss each measure using the same language. We therefore in this section introduce the dynamics of the model, both on the individual trajectory level and the ensemble level, as well as collect germane results regarding its energetics and thermodynamics.

2.1. Dynamics, energetics, and thermodynamics without feedback

Our quantity of interest is the time-dependent velocity v_t of a trapped, underdamped Brownian particle of mass m , coupled to a thermal reservoir at temperature T with viscous damping coefficient γ , evolving according to the Langevin equation [32]

$$m\dot{v}_t = -\gamma v_t + f_t + \xi_t, \quad (2)$$

where f_t is an externally controlled force, and ξ_t is zero-mean Gaussian white noise with covariance $\langle \xi_t \xi_s \rangle = 2\gamma T \delta(t - s)$. Starting here we set Boltzmann's constant to unity, $k_B = 1$.

In the absence of control, $f_t = 0$, the velocity v_t relaxes to an equilibrium Boltzmann distribution $p_{\text{eq}}(v) \propto \exp[-mv^2/(2T)]$. In the following, we will vary f_t using feedback

in order to cool the particle, that is damp its thermal fluctuations, thereby reducing its kinetic temperature $T_{\text{kin}} = m\langle v^2 \rangle < T$. Before we get to that, it is helpful to first review the energetics and thermodynamics of a driven, underdamped Brownian particle without feedback, so that we can appreciate the differences that arise in the presence of feedback.

To this end, we require the Fokker-Planck equation associated with (2) for the time-dependent probability density $p_t(v)$ [33],

$$\partial_t p_t(v) = -\partial_v J_t^v(v) \quad (3)$$

$$J_t^v(v) = -\frac{1}{m}(\gamma v_t - f_t)p_t(v) - \frac{\gamma T}{m^2}\partial_v p_t(v). \quad (4)$$

where we have introduced the (probability) current J_t^v . Anticipating our discussion of the thermodynamics, we divide the current into its irreversible half, which is anti-symmetric under time-reversal, and its reversible half, which is time-reversal symmetric, as [33, 34, 35]

$$J_t^{\text{irr}}(v) = -\frac{\gamma}{m}vp_t(v) - \frac{\gamma T}{m^2}\partial_v p_t(v) \quad (5)$$

$$J_t^{\text{rev}}(v) = \frac{f_t}{m}p_t(v). \quad (6)$$

Key to this splitting is treating the force f_t as even under time reversal, as typically assumed for a force arising from an external potential. With this identification, the irreversible portion of the current J_t^{irr} arises solely due to the forces imparted on the particle by its surroundings: the friction, $-\gamma v_t$, and the fluctuating force, ξ_t .

Moving on to the thermodynamics, we have from stochastic energetics an unambiguous identification of the heat flow into the system as the work done by the thermal reservoir on the particle [20, 34, 36, 37], which on average reads

$$\dot{Q} = \int mv J_t^{\text{irr}}(v)dv. \quad (7)$$

It notably only depends on the irreversible current arising from the forces due to the thermal reservoir. The particle's (internal) energy is its average kinetic energy

$$E = \left\langle \frac{1}{2}mv_t^2 \right\rangle = \int \frac{1}{2}mv^2 p_t(v)dv. \quad (8)$$

By differentiating E with time and substituting in the Fokker-Planck equation (3), we are able to identify the extracted work rate via the first law of thermodynamics $\dot{E} = -\dot{W}_{\text{ext}} + \dot{Q}$,

$$\dot{W}_{\text{ext}} = -\langle f_t v_t \rangle, \quad (9)$$

as the average power delivered against the external force f_t .

From stochastic thermodynamics, we also have the (irreversible) entropy production rate [34, 35, 37]

$$\dot{S}_i = \dot{S}(v) - \frac{\dot{Q}}{T} = \frac{m^2}{\gamma T} \int \frac{[J_t^{\text{irr}}(v)]^2}{p_t(v)} dv \geq 0, \quad (10)$$

where we have the traditional splitting into the time variation of the system's Shannon entropy $S(v) = - \int p_t(v) \ln p_t(v) dv$,

$$\dot{S}(v) = - \int J_t^v(v) \partial_v \ln p_t(v) dv, \quad (11)$$

and the reversible entropy exchange with the environment

$$\dot{S}_{\text{env}} = -\frac{\dot{Q}}{T}. \quad (12)$$

Notably, the entropy production only depends on the irreversible current, since it is a measure of the time-reversal symmetry breaking of the dynamics [34]. This property is what allowed us to pullout the contribution due to the heat, which is also only a function of the irreversible current.

2.2. Dynamics and energetics with feedback

Our main focus in this paper is feedback cooling, where we vary f_t in response to measurements of the velocity. Following [21], we consider a feedback protocol where we measure the velocity v_t obtaining outcomes y_t with some error, and then feed back those measurements by applying a force $f_t = -ay_t$ that acts as an additional friction, extracting work. A simple way to incorporate measurement error is to add to our read-out of v_t Gaussian white noise η_t of zero mean and covariance $\langle \eta_t \eta_s \rangle = \sigma^2 \delta(t-s)$, with σ^2 quantifying the measurement uncertainty: for example as $y_t = v_t + \eta_t$. However, white noise fluctuations are very violent. To make the problem more tractable, we smooth over the noise by applying a low-pass filter with time constant τ to the measurements: $y_t = (1/\tau) \int_0^t e^{-(t-s)/\tau} (v_s + \eta_s) ds$ [38]. We are therefore led to the following modified dynamics including measurement and feedback [21]

$$\begin{aligned} m\dot{v}_t &= -\gamma v_t - ay_t + \xi_t \\ \tau \dot{y}_t &= -(y_t - v_t - \eta_t), \end{aligned} \quad (13)$$

where a is the feedback gain. It is important to note at this point that y_t is merely a model of measurement outcomes. We are not making any assumption about the physical system that records the measurements, nor implements the feedback in response.

In general, the joint system relaxes to a time-independent, nonequilibrium steady state, where heat is continuously being extracted as work to maintain the particle at the cooled kinetic temperature. This is the scenario we focus on in the following.

To discuss the energetics, we need the equivalent description of the dynamics in (13) in terms of the Fokker-Planck equation for the time-dependent probability density $p_t(v, y)$,

$$\partial_t p_t(v, y) = -\partial_v J_t^v(v, y) - \partial_y J_t^y(v, y), \quad (14)$$

with (probability) currents

$$\begin{aligned} J_t^v(v, y) &= -\frac{1}{m}(\gamma v + ay)p_t(v, y) - \frac{\gamma T}{m^2} \partial_v p_t(v, y) \\ J_t^y(v, y) &= -\frac{1}{\tau}(y - v)p_t(v, y) - \frac{\sigma^2}{2\tau^2} \partial_y p_t(v, y). \end{aligned} \quad (15)$$

Again we can split the velocity current J_t^v into irreversible and reversible pieces, as in (5),

$$J_t^{\text{irr}}(v, y) = -\frac{\gamma}{m}vp_t(v, y) - \frac{\gamma T}{m^2}\partial_v p_t(v, y) \quad (16)$$

$$J_t^{\text{rev}}(v, y) = -\frac{a}{m}yp_t(v, y). \quad (17)$$

This splitting singles out the irreversible current as solely due to the thermal reservoir as before [cf. (5)], which is required to correctly link the heat and entropy production in the following. Again, this division relies on choosing $f_t = -ay_t$ as time-reversal symmetric, just as in the preceding section.

Our focus is the steady state solution, which due to the linear, Gaussian dynamics is the Gaussian probability density [32],

$$p_s(v, y) = \frac{1}{\sqrt{(2\pi)^2|\Sigma|}} \exp \left[-\frac{1}{2}(v, y) \cdot \Sigma^{-1} \cdot (v, y)^T \right], \quad (18)$$

where the steady-state covariance matrix is

$$\Sigma = \begin{pmatrix} \sigma_v^2 & \sigma_{vy} \\ \sigma_{vy} & \sigma_y^2 \end{pmatrix}, \quad (19)$$

and the associated steady-state currents are J_s^v and J_s^y . The entries of Σ can be determined by plugging (18) into (14), as detailed for a more general model in [21]; however their precise expressions are unilluminating and therefore relegated to Appendix A. We do observe that the reduced distribution of the velocity $p_s(v) = \int p_s(v, y) dy$ is also Gaussian. Therefore, it has the same structure as an equilibrium distribution, but with a smaller variance, or a cooler effective temperature [21]

$$T_{\text{kin}} = T \frac{1 + (a/\gamma)(a\sigma^2/(2T)) + (1 + a/\gamma)(\gamma\tau/m)}{1 + a/\gamma + (1 + a/\gamma)(\gamma\tau/m)} < T, \quad (20)$$

where the inequality is only satisfied in the regime of good cooling, $a\sigma^2 \leq 2T$. Otherwise too much measurement noise is fed back into the velocity, effectively heating it.

Again from stochastic energetics the heat current is identified as the energy lost due to the irreversible current arising from the thermal noise [20, 36, 37]

$$\dot{Q} = \int mvJ_t^{\text{irr}}(v, y) dv dy = \int mvJ_t^{\text{irr}}(v) dv, \quad (21)$$

which importantly only depends on the velocity as in (5), since the measurement and feedback do not affect the interaction with the thermal environment. In a similar way as before (9) the extracted work rate is

$$\dot{W}_{\text{ext}} = a\langle y_tv_t \rangle, \quad (22)$$

due to the correlations between the feedback force and the particle. In the steady state, \dot{W}_{ext} can be simplified using the defining equations for the elements of the covariance matrix Σ in Appendix A,

$$\dot{W}_{\text{ext}} = a\sigma_{vy} = \frac{1}{\tau_v} (T - T_{\text{kin}}), \quad (23)$$

in terms of the velocity's relaxation rate, $1/\tau_v = \gamma/m$. When the feedback is successful, and we have reduced the kinetic temperature $T_{\text{kin}} < T$, we must be extracting work, $\dot{W}_{\text{ext}} > 0$, recovering the results of [21].

We finally will require the fluctuating-trajectory solutions of (13) up to time t , $v_0^t = \{v_s\}_{s=0}^t$ and $y_0^t = \{y_s\}_{s=0}^t$. We can obtain the probability densities for these trajectories by discretizing time and then using the usual procedure for obtaining path-integral densities, which we sketch in Appendix B. The joint density $\mathcal{P}[v_0^t, y_0^t]$ can be conveniently expressed in terms of two probability densities

$$\hat{\mathcal{P}}[y_0^t|v_0^t, y_0] \propto \exp \left[- \int_0^t ds \frac{(\tau \dot{y}_s + y_s - v_s)^2}{2\sigma^2} \right], \quad (24)$$

suitably normalized, and

$$\hat{\mathcal{P}}[v_0^t|y_0^t, v_0] \propto \exp \left[- \int_0^t ds \frac{(m\dot{v}_s + \gamma v_s + ay_s)^2}{4\gamma T/m^2} \right], \quad (25)$$

as

$$\mathcal{P}[v_0^t, y_0^t] = \hat{\mathcal{P}}[v_0^t|y_0^t, v_0] \hat{\mathcal{P}}[y_0^t|v_0^t, y_0] p(v_0, y_0), \quad (26)$$

with initial probability density $p(v_0, y_0)$. It cannot be under emphasized that each $\hat{\mathcal{P}}$ is *not* the conditional probability of the feedback process, *i.e.*, $\hat{\mathcal{P}}[y_0^t|v_0^t, y_0] \neq \mathcal{P}[y_0^t|v_0^t, y_0] = \mathcal{P}[v_0^t, y_0^t|y_0]/\mathcal{P}[v_0^t|y_0]$, since v_t and y_t influence each other when there is feedback [8]. Instead, we can understand $\hat{\mathcal{P}}[y_0^t|v_0^t, y_0]$ by first imagining that we fix the *entire* velocity trajectory v_0^t , and then evolve y_t alone according to (13). This procedure has no feedback and the probability to observe a particular measurement trajectory is exactly $\hat{\mathcal{P}}[y_0^t|v_0^t, y_0]$. A similar interpretation holds for $\hat{\mathcal{P}}[v_0^t|y_0^t, v_0]$ as well. This distinction between $\hat{\mathcal{P}}$ and \mathcal{P} will become important in section 3.1 when we introduce the transfer entropy rate.

3. Information

In this section, we present the definitions of the various measures of information that can be used to bound the extracted work during a feedback process. In the next section, section 4, we will discuss the physics behind them.

3.1. Transfer entropy rate

The first information measure we discuss is the transfer entropy rate from v_t to y_t . The transfer entropy is a directional measure of information, which quantifies in an information-theoretic manner how much the dynamics (or more specifically the transition probabilities) of y_t are influenced by v_t [12]. For our continuous stochastic process, it reads

$$\dot{I}_{v \rightarrow y} = \lim_{t \rightarrow \infty} \frac{1}{t} \int \mathcal{D}[v_0^t] \mathcal{D}[y_0^t] \mathcal{P}[v_0^t, y_0^t] \ln \frac{\hat{\mathcal{P}}[y_0^t|v_0^t, y_0]}{\mathcal{P}[y_0^t|y_0]} \geq 0. \quad (27)$$

In Appendix B, we justify this expression by discretizing the evolution and then utilizing the well-developed theory for repeated, discrete feedback [3, 5, 7, 14, 25, 39]. When no measurements are taking place, the dynamics of y_t is independent of v_t , $\hat{\mathcal{P}}[y_0^t|v_0^t, y_0] = \mathcal{P}[y_0^t|y_0]$, and the transfer entropy rate is zero. On the other hand, the more influence the velocity has on the measurement outcomes the larger the transfer entropy rate. Furthermore, when there is only one measurement the transfer entropy simplifies to the mutual information [8]. An alternative, equivalent expression for the transfer entropy rate in the context of continuous feedback has been introduced by Sandberg *et al* [11]. A similar analysis was performed by Fujitani and Suzuki for discrete Markov processes [6, 25]. The transfer entropy rate in feedback systems described by continuous-time, discrete Markov processes has been extensively studied in [10, 14, 39, 40].

To compare $\dot{I}_{v \rightarrow y}$ with the other information measures, we calculate its value in our model of feedback cooling. The calculation is facilitated by noting that for stationary Gaussian processes, as we have, integrals of the form (27) can be conveniently expressed in terms of the power spectra – Fourier transforms of the correlation functions. For (27), we demonstrate in Appendix C that it can be formulated as

$$\dot{I}_{v \rightarrow y} = -\frac{1}{4\pi} \int_{-\infty}^{\infty} \ln \frac{\hat{C}_{yy|v}(\omega)}{C_{yy}(\omega)} d\omega, \quad (28)$$

where $C_{yy}(\omega)$ is the power spectrum of y_t , and $\hat{C}_{yy|v}(\omega)$ is the Fourier transform of the variance of y_t given a fixed trajectory v_0^t . We have carried out the integral in Appendix D with the result

$$\dot{I}_{v \rightarrow y} = \frac{\gamma}{2m} \left(\sqrt{1 + \frac{2T/\gamma}{\sigma^2}} - 1 \right) = \frac{1}{2\tau_v} \left(\sqrt{1 + \text{SNR}} - 1 \right). \quad (29)$$

New information is acquired at the relaxation rate of v_t , $\gamma/m = 1/\tau_v$; that is we learn new information about v_t only as fast as v_t changes enough to detect. In addition, the transfer entropy rate does not depend on the feedback parameters a and τ , but only on the measurement accuracy σ^2 through the dimensionless signal-to-noise ratio $\text{SNR} = (2T/\gamma)/\sigma^2$, which quantifies the relative size of the measurement accuracy to the thermal diffusion of the velocity. As a result, for perfect measurements without error, $\sigma = 0$, the SNR diverges and with it the transfer entropy rate. Thus, error-free measurement corresponds to infinite information, consistent with the notion that infinite information is required to localize a continuous variable with perfect precision.

3.2. Information flow

We next consider the information flow, whose origin is in the exchange of information between the velocity and the auxiliary measurement device implementing the control. It was first considered in the context of interacting diffusion processes [15], but subsequently has been introduced in the analysis of the thermodynamics of continuously-coupled, discrete stochastic systems [14, 16, 41]. When the coupling is not continuous,

but each system takes turns evolving, the information flow simplifies to the mutual information [16, 17, 18]. In order to facilitate connections to the other information measures, we sketch in this section the basic arguments leading to the information flow, following the program outlined in [16], and then calculate its value in our feedback cooling model.

First, we must note that strictly speaking this approach requires that y_t be the degree of freedom of a physical system, not simply an abstract measurement outcome. Still, in this section we would like not to comment on the precise thermodynamics of y_t , taking it only as a generic thermodynamic system. We will come back to its precise interpretation in section 4 when we compare the physics underlying the different information measures.

The key insight in this approach is that the (irreversible) entropy production of the joint system of v_t and y_t can be divided as

$$\dot{S}_i = \dot{S}_i^v + \dot{S}_i^y \geq 0, \quad (30)$$

with positive contributions arising due to the irreversible current in the v -direction (16),

$$\dot{S}_i^v = \frac{m^2}{\gamma T} \int \frac{[J_t^{\text{irr}}(v, y)]^2}{p_t(v, y)} dv dy \geq 0, \quad (31)$$

and separately from y_t , \dot{S}_i^y . The next step is to perform the traditional splitting of \dot{S}_i^v into the variation of the Shannon entropy due to v_t [cf. (11)],

$$\dot{S}(v) = - \int J_t^v(v, y) \partial_v \ln p_t(v) dv dy, \quad (32)$$

and the heat \dot{Q} (21) as

$$\dot{S}_i^v = \dot{S}(v) - \frac{\dot{Q}}{T} + \dot{I}_{\text{flow}} \geq 0. \quad (33)$$

The additional contribution due to the influence of y_t is an information-theoretic piece

$$\dot{I}_{\text{flow}} = - \int J_t^v(v, y) \partial_v \ln \frac{p_t(v, y)}{p_t(v)p_t(y)} dv dy, \quad (34)$$

which is (minus) the variation of the mutual information \ddagger

$$I(v_t; y_t) = \int p_t(v, y) \ln \frac{p_t(v, y)}{p_t(v)p_t(y)} dv dy \quad (35)$$

between v_t and y_t , due to the fluctuations of v_t [42]. The mutual information $I(v_t; y_t)$ is a measure of correlations, quantifying how knowledge of the measurement outcomes reduces uncertainty in the velocity. While \dot{I}_{flow} may be positive or negative, in the regime of good cooling where we are extracting work, we will always have $\dot{I}_{\text{flow}} \geq 0$. In the steady state, $J_s^v = 0$, and $\dot{Q} = \dot{W}_{\text{ext}}$, so that (33) reduces to [14, 16]

$$\dot{W}_{\text{ext}} \leq T \dot{I}_{\text{flow}}, \quad (36)$$

\ddagger We have defined the information flow with the opposite sign convention of [14, 15, 16, 41], so that it is positive in the cooling regime, allowing a straightforward comparison to the other information measures.

in the form of (1).

Employing the steady-state solution in (18), we have for the steady-state information flow

$$\dot{I}_{\text{flow}} = \frac{\gamma}{m} \left(\frac{T}{m} \frac{\sigma_y^2}{|\Sigma|} - 1 \right) \geq 0. \quad (37)$$

where $|\Sigma|$ denotes the determinant of Σ . Unfortunately, we have been unable to formulate a more transparent expression in general. Even still, the information rate again only grows as fast as the relaxation rate of the velocity $\gamma/m = 1/\tau_v$.

3.3. Entropy pumping

For the feedback cooling of a Brownian particle without errors an entropy pumping bound has been introduced by Kim and Qian [19]. This approach has subsequently been developed by Ge [43] and extended to the setup in (13) by Munakata and Rosinberg [20, 21, 22], which we discuss in this section.

The entropy pumping approach is based on a coarse graining of the Fokker-Planck equation (14). Following [21], we formally integrate out y_t from (14) to obtain the reduced Fokker-Planck equation

$$\partial_t p_t(v) = \partial_v \left(\frac{1}{m} \left(\gamma v + \tilde{f}_t^{\text{fb}}(v) \right) p_t(v) + \frac{\gamma T}{m^2} \partial_v p_t(v) \right) = -\partial_v \tilde{J}_t(v), \quad (38)$$

where we have identified an effective feedback force

$$\tilde{f}_t^{\text{fb}}(v) = a \int y p_t(y|v) dy. \quad (39)$$

Furthermore, we treat \tilde{f}_t^{fb} as time-reversal symmetric, as we would expect for an external force [21]. In which case, we single out from the coarse-grained current the irreversible current exactly as for the no-feedback case (5),

$$\tilde{J}_t(v) = J_t^{\text{irr}}(v) - \frac{1}{m} \tilde{f}_t^{\text{fb}}(v). \quad (40)$$

This will allow us to connect the entropy production in the environment with the heat.

Equation (38) is not a closed equation for $p_t(v)$; the measurement dynamics are required to solve it. Nevertheless, the entropy pumping approach is to treat (38) as a thermodynamically consistent equation for $p_t(v)$ with an effective external force \tilde{f}_t^{fb} . In this case, the entropy balance is developed in analogy to the no-feedback setup, as in (10),

$$\dot{S}_i^v = \frac{m^2}{\gamma T} \int \frac{[J_t^{\text{irr}}(v)]^2}{p_t(v)} dv = \dot{S}(v) - \frac{\dot{Q}}{T} + \dot{I}_{\text{pump}} \geq 0, \quad (41)$$

where the second equality follows by substituting in definition of the coarse-grained current $\tilde{J}_t(v)$ in (40). Here, $\dot{S}(v) = -\int \tilde{J}_t(v) \partial_v \ln p_t(v) dv$ is equivalent to the expression for the rate of change of the system's Shannon entropy including feedback in (32), and the additional entropy pumping term arises due to the coarse-grained feedback force,

$$\dot{I}_{\text{pump}} = \int p_t(v) \partial_v \frac{1}{m} \tilde{f}_t^{\text{fb}}(v) dv. \quad (42)$$

As pointed out in [21], the feedback force is proportional to the minimum mean square error estimate of y_t given v_t . Other than that though, there does not appear to be a crisp interpretation of the entropy pumping as a form of information, like for the transfer entropy rate and information flow.

Using the steady-state distribution in (18), we have for the steady-state entropy pumping [21]

$$\dot{I}_{\text{pump}} = \int p_s(v) \partial_v \frac{1}{m} \tilde{f}_s^{\text{fb}}(v) dv = \frac{a}{m} \frac{\sigma_{vy}}{\sigma_v^2} = \frac{1}{\tau_v} \left(\frac{T - T_{\text{kin}}}{T_{\text{kin}}} \right) \geq 0, \quad (43)$$

with positivity guaranteed when there is cooling $T \geq T_{\text{kin}}$.

3.4. Trajectory mutual information

Another information measure that has aroused some attention is the mutual information rate between the entire v_0^t and y_0^t trajectories [13, 44]. For continuous stochastic processes, the trajectory mutual information rate is [42]

$$\dot{I}_{\text{traj}} = \lim_{t \rightarrow \infty} \frac{1}{t} I(v_0^t; y_0^t) = \lim_{t \rightarrow \infty} \frac{1}{t} \int \mathcal{D}[v_0^t] \mathcal{D}[y_0^t] \mathcal{P}[v_0^t, y_0^t] \ln \frac{\mathcal{P}[v_0^t, y_0^t]}{\mathcal{P}[v_0^t] \mathcal{P}[y_0^t]} \geq 0. \quad (44)$$

It quantifies how much the uncertainty about the entire velocity trajectory v_0^t is reduced given knowledge of the entire measurement trajectory y_0^t , and vice versa, as it is symmetric.

The \dot{I}_{traj} bound on the extracted work follows readily once we observe a close connection between the trajectory mutual information and the transfer entropy pointed out in [14]; by substituting \mathcal{P} with $\hat{\mathcal{P}}$ (26) in \dot{I}_{traj} , it follows that

$$\dot{I}_{\text{traj}} = \dot{I}_{v \rightarrow y} + \dot{I}_{y \rightarrow v}, \quad (45)$$

after identifying the transfer entropy rate from y_t to v_t , $\dot{I}_{y \rightarrow v} \geq 0$, defined analogously to $\dot{I}_{v \rightarrow y}$ (27). The positivity of the transfer entropy implies that

$$\dot{W}_{\text{ext}} \leq T \dot{I}_{v \rightarrow y} \leq T \dot{I}_{\text{traj}}, \quad (46)$$

giving (1) for the trajectory information, which is always weaker than the transfer entropy bound.

The trajectory information rate has been studied in numerous contexts and has a well-known expression in terms of power spectra [45, 46, 47, 48] that we recall in Appendix C,

$$\dot{I}_{\text{traj}} = -\frac{1}{4\pi} \int_{-\infty}^{\infty} \ln \left(1 - \frac{|C_{vy}(\omega)|^2}{C_{vv}(\omega)C_{yy}(\omega)} \right) d\omega. \quad (47)$$

In Appendix D, we perform this integral to find

$$\dot{I}_{\text{traj}} = \frac{\gamma}{2m} \left(\sqrt{1 + \frac{2T/\gamma}{\sigma^2}} - 1 \right) + \frac{1}{2\tau} \left(\sqrt{1 + \frac{a^2\sigma^2}{2\gamma T}} - 1 \right). \quad (48)$$

Comparing with (45), we have as a byproduct the transfer entropy rate from y_t to v_t ,

$$\dot{I}_{y \rightarrow v} = \frac{1}{2\tau} \left(\sqrt{1 + \frac{a^2\sigma^2}{2\gamma T}} - 1 \right). \quad (49)$$

3.5. Maximum work

A final bound on the extracted work is simply to maximize \dot{W}_{ext} in (23) with respect to the measurement parameters a and τ . While the result is not general, remarkably for linear Gaussian processes it has a close connection with the transfer entropy rate, as first noticed by Sandberg *et al* [11]. Using standard calculus methods, the extracted work is bounded above by its maximal value

$$\dot{W}_{\text{ext}} \leq \dot{W}_{\text{ext}}^{\text{max}} = T_{\text{kin}} \dot{I}_{v \rightarrow y}, \quad (50)$$

akin to (1), for parameter values

$$a^* = \gamma \left(\sqrt{1 + \frac{2T/\gamma}{\sigma^2}} - 1 \right), \quad \tau^* = 0. \quad (51)$$

The optimal measurement has no low-pass filtering: It is immediately fed back into the particle to control it. Remarkably, the extracted work is again bounded by the transfer entropy rate, except multiplied by the cooled kinetic temperature of the particle, instead of T .

3.6. Discussion

3.6.1. Quantitative comparison of information measures. To better understand the relationship between all of these information measures, we plot them all together with \dot{W}_{ext} in figure 1 as a function of the feedback gain a and measurement error σ^2 in the range where cooling occurs ($\dot{W}_{\text{ext}} \geq 0$). As expected, each information measure bounds the extracted work.

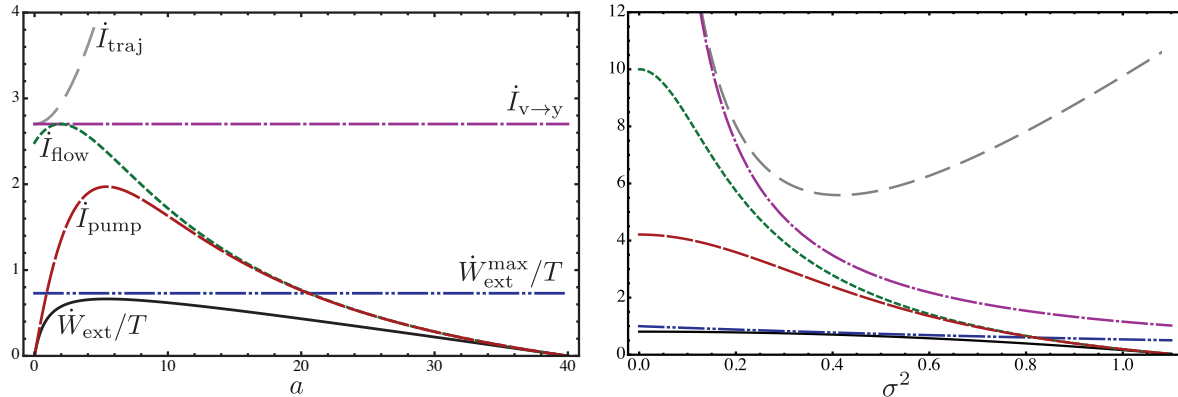


Figure 1. Comparison of the information measures and the extracted work \dot{W}_{ext}/T in dimensionless units as a function of the feedback gain a (left) and measurement error σ^2 (right). Parameters $m = 1$, $\gamma = 1$, $T = 5$, $\tau = 0.1$, with $\sigma = 0.5$ (left) and $a = 8$ (right)

The most striking feature of figure 1 is the hierarchy of information measures,

$$\dot{I}_{\text{traj}} \geq \dot{I}_{v \rightarrow y} \geq \dot{I}_{\text{flow}} \geq \dot{I}_{\text{pump}}, \quad (52)$$

apart from $\dot{W}_{\text{ext}}^{\text{max}}$, which does not actually have a generic information interpretation. In fact, this ranking holds quite generally. We have already seen that $\dot{I}_{\text{traj}} \geq \dot{I}_{v \rightarrow y}$ in section 3.4 when discussing the second-law-like inequality for the trajectory information. The middle inequality, $\dot{I}_{v \rightarrow y} \geq \dot{I}_{\text{flow}}$, has been demonstrated by Hartich *et al* [14] for continuous-time, discrete Markov jump processes. For diffusion processes, a similar conclusion was reached by Allahverdyan *et al* [15] except for a slightly different transfer entropy rate that uses only the most recent measurement, which upper bounds the transfer entropy rate considered here, as pointed out in [14]. Nevertheless, the proof for jump processes in [14] can be carried over to diffusion processes, once their evolution is discretized. The last inequality between the information flow and the entropy pumping also is generic. This follows by bounding the steady-state entropy production of v_t in the information-flow description (31) using a coarse-graining inequality [21] to connect it to the coarse-grained, entropy-pumping approach:

$$\dot{S}_i^v = -\frac{\dot{Q}}{T} + \dot{I}_{\text{flow}} = \frac{m^2}{\gamma T} \int \frac{[J_s^{\text{irr}}(v, y)]^2}{p_s(v, y)} dv dy \geq \frac{m^2}{\gamma T} \int \frac{[J_s^{\text{irr}}(v)]^2}{p_s(v)} dv = -\frac{\dot{Q}}{T} + \dot{I}_{\text{pump}} = \dot{S}_i^v \quad (53)$$

where we have employed the entropy balance of entropy pumping in (41). Clearly,

$$\dot{I}_{\text{flow}} \geq \dot{I}_{\text{pump}}. \quad (54)$$

As a lower bound on all other information measures, the entropy pumping can be given an information-theoretic interpretation, which till now has been lacking, as a minimal information requirement for successful feedback cooling.

An alternative perspective on this hierarchy of information measures comes from considering the efficiency of work extraction

$$\varepsilon = \frac{\dot{W}_{\text{ext}}}{T\dot{I}} \leq 1. \quad (55)$$

By utilizing the smaller information measures, we will estimate higher efficiencies, even without changing the measurement or feedback procedure. This conclusion is somewhat surprising, since it makes the notion of efficiency somewhat arbitrary. We will come back to this observation later, after discussing the physical origins of the different information measures.

We also see in figure 1 that the transfer entropy rate and the trajectory mutual information diverge as the measurement error tends to zero, $\sigma^2 \rightarrow 0$; whereas the other measures remain finite. Munakata and Rosinberg have also observed that the entropy pumping displays a nontrivial structure, attaining a maximum at the maximum cooling rate [21]. Figure 1 demonstrates that \dot{I}_{flow} displays a similar structure, but its maximum does not quite correspond to the maximum cooling. Most likely, this discrepancy arises due to the effect of coarse-graining.

3.6.2. Optimal control and the Kalman-Bucy filter. Interestingly, closer connections exist between the information flow, transfer entropy rate, and maximum extractable work that are revealed by re-examining our feedback problem from the perspective of optimal control theory.

The feedback cooling we have been addressing is a special case of a classic problem in optimal control theory: the characterization of feedback controllers that minimize quadratic performance objectives of the form

$$\mathcal{J} = \langle v_t^2 \rangle + \rho \langle f_t^2 \rangle, \quad (56)$$

where $\rho > 0$ is a constant parameter used to tune the trade-off between keeping small fluctuations in v_t and applying a strong control force f_t , for example [49]. For the special case of cooling, we have been focused on minimizing $\langle v_t^2 \rangle$ alone, which corresponds to $\rho \rightarrow 0$.

Assuming linear dynamics and Gaussian noise, the optimal feedback controller with access to noisy measurements $v_t + \eta_t$ can be written in the form

$$\begin{aligned} m\dot{\hat{v}}_t &= -\gamma\hat{v}_t - G\hat{v}_t + K(v_t + \eta_t - \hat{v}_t) \\ f_t &= -G\hat{v}_t, \end{aligned} \quad (57)$$

where \hat{v}_t is the abstract dynamical state of the controller, and G and K are carefully chosen constants. According to the separation principle [49, 50], these parameters G and K can be determined as the solutions to two independent optimization problems: the optimal gain G is obtained by minimizing \mathcal{J} , temporarily assuming there is no measurement noise, $\sigma = 0$; whereas the optimal K is obtained by minimizing the estimation error, see below, and is independent of the tuning parameter ρ . While the exact expression for the optimal gain G is of little interest to us here, we do note that it tends monotonically to infinity as $\rho \rightarrow 0$. This makes intuitive sense, since $\rho \rightarrow 0$ means we only care about minimizing the variance $\langle v_t^2 \rangle$ and assess no cost for large control forces $\langle f_t^2 \rangle$. On the other hand, optimal filtering theory selects an optimal K by minimizing the estimation error,

$$\mathcal{E}_t \equiv \min_K \langle (v_t - \hat{v}_t)^2 \rangle, \quad (58)$$

given all the past measurements $(v + \eta)_0^t$. The steady-state optimum, achieved for

$$K = \gamma \left(\sqrt{1 + \frac{2T/\gamma}{\sigma^2}} - 1 \right), \quad (59)$$

is

$$\mathcal{E} = \frac{\sigma^2}{m} K = \sigma^2 \frac{\gamma}{m} \left(\sqrt{1 + \frac{2T/\gamma}{\sigma^2}} - 1 \right). \quad (60)$$

Thus, \hat{v}_t represents the best estimate of v_t given all past measurements. In fact, no other filter, nonlinear or otherwise, can produce a better estimate than the one described here, which is known as the Kalman-Bucy filter [49, 51].

Remarkably, the optimal controller (57) with Kalman-Bucy filter can always be realized using the feedback cooling dynamics in (13) by a simple rescaling

$$\hat{v}_t = \left(\frac{K}{\gamma + K + G} \right) y_t, \quad (61)$$

and choosing the parameters a and τ as

$$a_{\text{KB}} = \frac{GK}{\gamma + K + G}, \quad \tau_{\text{KB}} = \frac{m}{\gamma + K + G}. \quad (62)$$

This mapping allows us to investigate our information measures from a new point of view by replacing y_t with the optimal \hat{v}_t .

For starters, maximal cooling, which coincides with the maximum extracted work $\dot{W}_{\text{ext}}^{\text{max}}$ (50), is obtained when $G \rightarrow \infty$, in which case the optimal controller (62) becomes

$$a_{\text{KB}} = K = \gamma \left(\sqrt{1 + \frac{2T/\gamma}{\sigma^2}} - 1 \right), \quad \tau_{\text{KB}} = 0, \quad (63)$$

recovering a^* and τ^* in (51) as expected.

The optimal controller also extracts the maximum amount of information. To see this, first note that optimality of the estimate \hat{v}_t implies that the estimation error is stochastically orthogonal to the estimate $\langle \hat{v}_t(v_t - \hat{v}_t) \rangle = 0$ for all t [49]. This property greatly simplifies the steady-state covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_v^2 & \sigma_{v\hat{v}} \\ \sigma_{v\hat{v}} & \sigma_{\hat{v}}^2 \end{pmatrix} = \begin{pmatrix} \sigma_{\hat{v}}^2 + \mathcal{E} & \sigma_{\hat{v}}^2 \\ \sigma_{\hat{v}}^2 & \sigma_{\hat{v}}^2 \end{pmatrix}, \quad (64)$$

where the variance of the estimate is simply

$$\sigma_{\hat{v}}^2 = \frac{K^2 \sigma^2}{2m(\gamma + G)}. \quad (65)$$

Note optimal cooling is achieved by $G \rightarrow \infty$, forcing $\sigma_{\hat{v}}^2 \rightarrow 0$, so that fluctuations in the velocity $\sigma_v^2 = \mathcal{E}$ are only caused by estimation error. Furthermore, by exploiting the structure of Σ in (64), the expression for the steady-state information flow (37) greatly simplifies,

$$\dot{I}_{\text{flow}} = \frac{\gamma}{m} \left(\frac{T}{m} \frac{\sigma_{\hat{v}}^2}{|\Sigma|} - 1 \right) = \frac{\gamma}{m} \left(\frac{T}{m\mathcal{E}} - 1 \right) = \frac{\gamma}{2m} \left(\sqrt{1 + \frac{2T/\gamma}{\sigma^2}} - 1 \right) = \dot{I}_{v \rightarrow y}, \quad (66)$$

for all G . This is a very interesting observation, supporting the claimed optimality of the Kalman-Bucy filter. We already know that $\dot{I}_{\text{flow}} \leq \dot{I}_{v \rightarrow y}$. What we see here is that the class of controllers given by (57), *i.e.*, with K fixed (59) and G free, saturates the bound, maximizing the information flow. Hence, a controller with a small gain G (zero even) only uses information to create an optimal estimate of the process, whereas a high gain cools as well. To gain further insight into equality (66), we have to look at the transfer entropy rate and information flow from a different perspective. Namely, the transfer entropy rate can also be defined as the rate of growth of the mutual information between v_t and the entire trajectory of measurement outcomes y_0^t , that is the change in $I(v_t; y_0^t)$. On the other hand, the information flow is the rate of growth of the mutual information between v_t and just the most recent measurement y_t , that is the change in $I(v_t; y_t)$. The inequality $\dot{I}_{\text{flow}} \leq \dot{I}_{\text{trans}}$ is then related to the simple idea that the entire trajectory of measurements contains more information than just the last. Now, it is known that the Kalman-Bucy filter \hat{v}_t is a sufficient statistic for the conditional

distribution of v_t given the measurements y_0^t [52]. In other words, everything useful in a collection of measurements for predicting v_t is contained in just \hat{v}_t , or in terms of the mutual information $I(v_t; \hat{v}_0^t) = I(v_t; \hat{v}_t)$. This equality translated into rates implies (66).

In figure 2, we illustrate how the extracted work depends on G , and how the maximum is asymptotically achieved. In addition, we see that $\dot{I}_{\text{flow}} = \dot{I}_{v \rightarrow y}$ holds for all G . We can also conclude that with certain choices of a and τ (namely a_{KB} and τ_{KB} in (62)) our original setup (13) can always saturate $\dot{I}_{\text{flow}} \leq \dot{I}_{v \rightarrow y}$, which is indeed observed in figure 1 for $a \approx 2$.

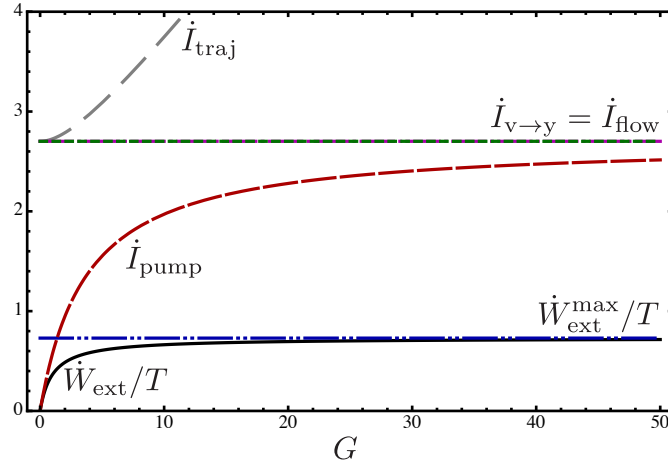


Figure 2. Comparison of the information measures for the Kalman-Bucy filter with the extracted work \dot{W}_{ext}/T in dimensionless units as a function of the Kalman-Bucy gain G with parameters $m = 1$, $\gamma = 1$, $T = 5$, $\sigma = 0.5$, and $\tau = 0.1$.

4. Energetics of Information and Measurement

We have seen that there are various, distinct measures of information that each offer a nontrivial bound for the extracted work. However, there does not seem to be an obvious reason to prefer any of one these measures. To this end, we investigate their origins in this section. We will find that the transfer entropy rate and the information flow both correspond to the information that is recorded in an auxiliary system, or memory, and therefore is subject to the limits of thermodynamics, as originally suggested by Landauer [2]. In particular, we show that these two information measures both bound the minimum energy required to gather that information through distinct thermodynamic processes, implying that the energy that we are able to extract as work originates in the (free) energy supplied by the memory.

4.1. Information flow

Let us start with the simpler measurement scenario corresponding to the information flow \dot{I}_{flow} . Actually, we have already touched on its physical interpretation when we

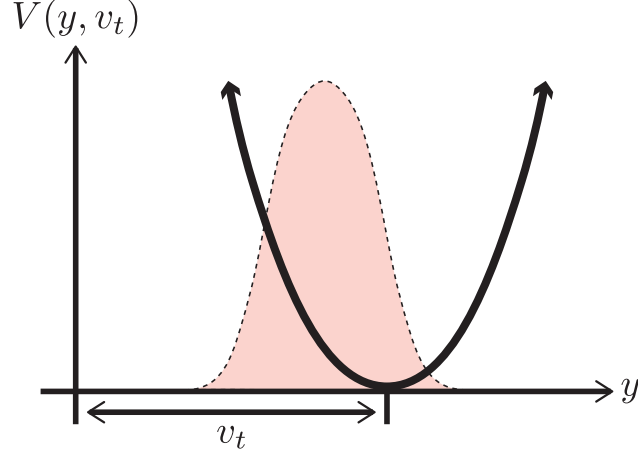


Figure 3. Illustration of the moving potential experienced by the memory degree of freedom, $V(y, v_t) = (y - v_t)^2/2$, centered about the time-dependent velocity v_t . The instantaneous probability density of y_t (shaded pink region) is Gaussian and lags behind the potential due to the finite relaxation time τ .

introduced it in section 3.2. Recall, there we considered the measurement outcomes y_t to correspond to a physical degree of freedom of an auxiliary system. We now clarify that interpretation by taking y_t to be the position of a secondary, harmonically-trapped, *overdamped* Brownian particle. To be thermodynamically consistent, the origin of the measurement noise must be a thermal reservoir, which requires imposing the Fluctuation-Dissipation theorem [37]:

$$\sigma^2 = 2\tau T. \quad (67)$$

We have chosen the temperature of the measurement device to be the same as the controlled system, which is the customary choice. From this point of view, (13) is the equation of motion for an overdamped Brownian particle with viscous damping coefficient τ trapped in a harmonic potential $V(y, v) = (y - v)^2/2$ of unit spring constant, centered about the velocity, as illustrated in figure 3. Alternatively, such a coupling can be implemented in an electric circuit as was presented in [11]. The result is that the position of the measurement oscillator y_t feels a fluctuating force making it track the velocity v_t , thereby establishing and maintaining correlations. Roughly speaking, the measurement oscillator is constantly learning new information at a rate \dot{I}_{flow} , which keeps getting rewritten in the value of its position.

When introducing the information flow, we divided the entropy production into two positive contributions (30), one due to the velocity \dot{S}_i^v , and another due to the measurements \dot{S}_i^y . When studying the extracted work \dot{W}_{ext} , we focused on \dot{S}_i^v . However, a similar analysis also holds for \dot{S}_i^y , which verifies that the y -system must consume at least \dot{I}_{flow} free energy to sustain the correlations that promote feedback. Observing that as a position y_t is even under time-reversal (consistent with our previous analysis in section 2.1), we develop its thermodynamics by splitting its current J_t^y (15) into

irreversible and reversible portions as

$$\begin{aligned} J_t^{\text{irr},y}(v, y) &= -\frac{1}{\tau} y p_t(v, y) - \frac{T}{\tau} \partial_y p_t(v, y) \\ J_t^{\text{rev},y}(v, y) &= \frac{v}{\tau} p_t(v, y) \end{aligned} \quad (68)$$

Notice that here the irreversible current is the time-symmetric contribution, since y_t is even under time-reversal [34]. Then, repeating the analysis in section 3.2, we have that in the steady state [14, 15, 16]

$$\dot{S}_i^y = \frac{\tau}{T} \int \frac{[J_s^{\text{irr},y}(v, y)]^2}{p_s(v, y)} dv dy = \frac{\dot{Q}^y}{T} - \dot{I}_{\text{flow}} \geq 0, \quad (69)$$

where $\dot{Q}^y = -\int y J_s^{\text{irr},y}(v, y) dv dy = (\sigma_y^2 - T)/\tau$ is the heat flow rate into y 's reservoir. Thus, in the steady state

$$\dot{Q}^y \geq T \dot{I}_{\text{flow}}. \quad (70)$$

In order to track the velocity, y 's environment continually absorbs heat at a rate \dot{Q}^y . We verify this bound in figure 4, where \dot{Q}^y is plotted with \dot{I}_{flow} . The minimum

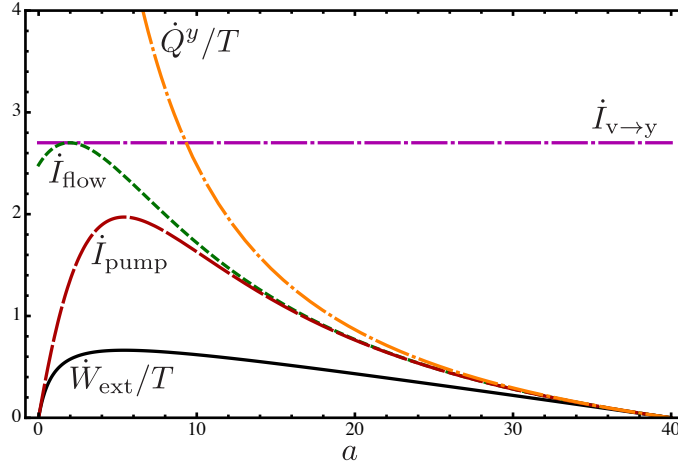


Figure 4. Plot of the heat dissipated by the auxiliary measurement oscillator \dot{Q}^y as a function of the gain a demonstrating that it upper bounds the information flow \dot{I}_{flow} , entropy pumping \dot{I}_{pump} , and extracted work \dot{W}_{ext} . The transfer entropy rate $\dot{I}_{v \rightarrow y}$ shares no relation with \dot{Q}^y . Parameters are $m = 1$, $\gamma = 1$, $T = 5$, $\sigma = 0.5$, and $\tau = 0.1$.

$\dot{Q}^y = T \dot{I}_{\text{flow}}$ is reached when the measurement device operates reversibly in the limit $\tau \ll \tau_v$, so that y_t rapidly relaxes to its instantaneous equilibrium centered about v_t : $p_s(y|v) \propto \exp[-(y - v)^2/(2T)]$.

In addition, we have already argued that the entropy pumping lower bounds the information flow, $\dot{I}_{\text{flow}} \geq \dot{I}_{\text{pump}}$ (54). As a result, \dot{I}_{pump} offers a weaker lower bound on the energy required for an auxiliary system to provide the entropy-pumping feedback, $\dot{Q}^y \geq T \dot{I}_{\text{pump}}$, which is verified in figure 4 as well.

4.2. Transfer entropy rate

The transfer entropy rate can also be understood as the minimum free energy required to measure, but with an alternative measurement scenario. In the previous section, the information flow was shown to bound the energy required to constantly rewrite a *single* memory with each new measurement y_t . By contrast, the setup for the transfer entropy rate is much closer to that envisioned by Landauer and Bennett in their thermodynamics of computation [2, 53]: Each measurement is recorded separately in one of a *collection* of memories through a specific driven thermodynamic process; one example of which was recently described in [23].

The central idea is that each measurement outcome is recorded in a distinct memory. Therefore, to track the system over any finite time interval, say from time $s = 0$ to t , we require an infinite number of memories in which to record the infinity of measurements. However, this is difficult to analyze. So to proceed, we discretize time as $s_k = k\Delta s$, with $k = 0, \dots, N$ and $\Delta s = t/N$, where the measurement outcome at time s_k is denoted simply as $y_k \equiv y_{s_k}$, and similarly $v_k \equiv v_{s_k}$. To store these measurement outcomes, we imagine a collection of N auxiliary memories with phase space positions m_k , prepared initially in positions $m_{k,0}$ distributed according to $\rho_0(m_{k,0})$. The measurement is a thermodynamic process during a time interval of length θ in which the k -th memory is manipulated, with the velocity fixed, in such a way to reproduce the correlations with v_{k-1} embodied in the measurement statistics of y_k . In other words, we demand that the statistics of the k -th memory after the measurement are $m_{k,\theta} \sim y_k$ (equality in distribution).

To see how these ideas play out in our model system, consider the discretized version of the Langevin equation (13)

$$y_k = y_{k-1} - \frac{\Delta s}{\tau}(y_{k-1} - v_{k-1}) + \frac{1}{\tau}\Delta\eta_{k-1}, \quad (71)$$

where the $\Delta\eta_k$ are independent Gaussian random variables of zero mean and covariance $\langle\Delta\eta_k\Delta\eta_l\rangle = \sigma^2\Delta s\delta_{kl}$. Equation (71) is a rule that tells us how the measurement outcome y_k at time s_k depends on the velocity v_{k-1} as well as the past measurement outcome y_{k-1} stored in a previous memory. Such measurements that depend on past outcomes are sometimes called non-Markovian measurements [40]. Specifically, y_k is characterized by a Gaussian probability density

$$P(y_k|y_{k-1}, v_{k-1}) = \frac{1}{\sqrt{2\pi\Omega^2}} \exp\left\{-\frac{[y_k - \mu_k(v_{k-1}, y_{k-1})]^2}{2\Omega^2}\right\} \quad (72)$$

$$\mu_k(v_{k-1}, y_{k-1}) = y_{k-1} - \frac{\Delta s}{\tau}(y_{k-1} - v_{k-1}), \quad \Omega^2 = \frac{\sigma^2\Delta s}{\tau^2},$$

roughly centered about the velocity with a variance depending on the measurement error. Now, in view of our previous discussion, we desire a physical system to act as a memory and a measurement process that prepares that system in a statistical state with the probability density in (72). A natural choice is an overdamped harmonic oscillator coupled to a thermal reservoir at temperature T . Initially each memory oscillator

is prepared in equilibrium with an arbitrary initial spring constant k_0 centered about zero, as illustrated in figure 5. Since each measurement is performed in sequence, it is

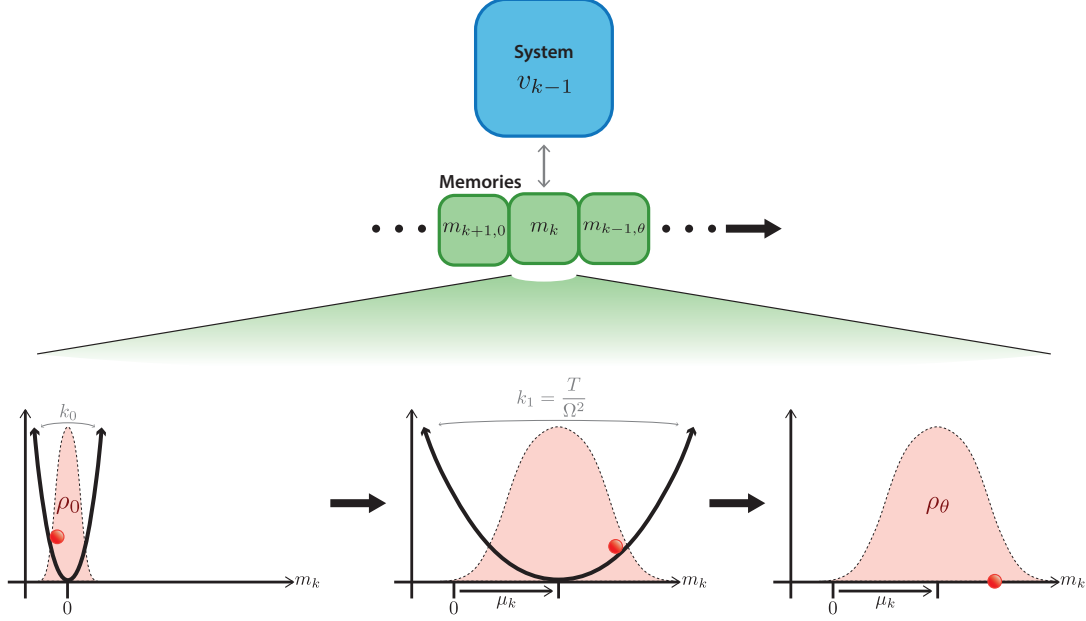


Figure 5. Schematic illustration of the transfer entropy rate measurement scenario: At time s_k , the velocity v_{k-1} is recorded in the k -th memory, harmonic oscillator (red dot) with initial state $m_{k,0}$ through a nonautonomous interaction that slowly shifts and expands its potential $V(m_k, v_{k-1}, m_{k-1})$, before quickly turning off. Concurrently, the probability density (pink shaded region) expands from $\rho_0(m_{k,0})$ to a width Ω^2 and shifts by $\mu_k(v_{k-1}, y_{k-1})$, terminating the process in the measurement probability density $\rho_\theta(m_{k,\theta})$ equivalent to (72), correlated with v_{k-1} and the past measurement outcome y_{k-1} stored in the previous memory state $m_{k-1,\theta}$. The process is then repeated, with each new measurement recorded in the next memory the tape.

attractive to visualize the phase spaces of the N measurement oscillators aligned in a row, or tape. Then one by one we couple each measurement oscillator to the system as well as past memories, so as to establish correlations. The density in (72) suggests that the measurement protocol for the k -th oscillator should be the quasistatic turn-on of an interaction that shifts the center of the harmonic oscillator to μ_k – which includes interactions with the past memories – while simultaneously expanding the spring constant to $k_1 = T/\Omega^2$, which results in the interaction potential

$$V(m_k, m_{k-1}, v_{k-1}) = \frac{T}{2\Omega^2} [m_k - \mu_k(v_{k-1}, m_{k-1})]^2 \quad (73)$$

as depicted in figure 5. As a result, upon completion of the k -th measurement the memory's position $m_{k,\theta}$ has settled into an equilibrium distribution $\rho_\theta(m_{k,\theta}|m_{k-1,\theta}, v_{k-1}) \propto \exp[-V(m_{k,\theta}, m_{k-1,\theta}, v_{k-1})/T]$ equivalent to (72). To complete the measurement, we must freeze the state of the memory to lock in the correlations, and remove the interactions by turning off V . One possible, though admittedly idealized, option is to

instantaneously set $V = 0$, and then immediately turn off the dynamics of the measurement oscillator – perhaps by quenching the temperature to zero – so that the oscillator can no longer move. By repeating this sequence of actions on each new memory, we store a collection of measurement outcomes, each in a different physical memory. Now to be precise each measurement has to be performed instantaneously from the point of view of the velocity. This merely means that the time-scale of the evolution of the individual memories must be much faster than the velocity, $\theta \ll \tau_v$, so that the measurement is completed before the velocity changes appreciably [23]. However, this assumption is not unreasonable, since measurements are usually assumed to read out the instantaneous state of the system.

Having described how we can mimic the measurement statistics in a physical situation, we now address the thermodynamics from a general point of view, applying the methods of [17, 18, 23]. Our analysis is based on the following second-law-like inequality that relates the work performed in an isothermal process to the increment in the nonequilibrium free energy [54, 55]: For a thermodynamic system with microscopic states z , the work W performed along an isothermal process during which the system's probability density transitions from $p(z)$ to $p'(z')$ is bounded as

$$W - \Delta\mathcal{F}(z') \geq 0. \quad (74)$$

where $\Delta\mathcal{F}(z') = \mathcal{F}(z') - \mathcal{F}(z)$ is the change in the nonequilibrium free energy $\mathcal{F}(z) = U(z) - TS(z)$ defined in terms of the average internal energy $U(z)$ and entropy $S(z) = -\int p(z) \ln p(z) dz$. The nonequilibrium free energy is a natural extension of the equilibrium free energy to systems characterized by an arbitrary probability density, since it reduces to the equilibrium free energy for systems in equilibrium.

We begin by focusing on the work done during the k -th measurement, W_k , during which the k -th memory becomes correlated with not only the velocity v_{k-1} but all the past memories $m_0^{k-1} = \{m_{l,\theta}\}_{l=0}^{k-1} \sim \{y_l\}_{l=0}^{k-1}$ through the velocity which depends on the entire past. Applying (74), we have

$$W_k - \Delta\mathcal{F}(m_{k,\theta}|m_0^{k-1}, v_{k-1}) \geq 0, \quad (75)$$

where $\Delta\mathcal{F}(m_{k,\theta}|m_0^{k-1}, v_{k-1}) = \mathcal{F}(m_{k,\theta}|m_0^{k-1}, v_{k-1}) - \mathcal{F}(m_{k,0}|m_0^{k-1}, v_{k-1})$ is the change in the nonequilibrium free energy of the k -th memory, corresponding to the change in the conditional density from $\rho_0(m_{k,0}|m_0^{k-1}, v_{k-1}) = \rho_0(m_{k,0})$ – due to the initial independence of each memory – to $\rho_\theta(m_{k,\theta}|m_0^{k-1}, v_{k-1}) = \rho_\theta(m_{k,\theta}|m_{k-1,\theta}, v_{k-1})$. We single out the new correlations by introducing the mutual information between $m_{k,\theta}$ and v_{k-1} conditioned on all the past measurements as $I(m_{k,\theta}; v_{k-1}|m_0^{k-1}) = S(m_{k,\theta}|m_0^{k-1}) - S(m_{k,\theta}|m_0^{k-1}, v_{k-1})$ [42]. Substituting in this definition, (75) becomes

$$W_k - \Delta\mathcal{F}(m_{k,\theta}|m_0^{k-1}) \geq TI(m_{k,\theta}; v_{k-1}|m_0^{k-1}), \quad (76)$$

where $\Delta\mathcal{F}(m_{k,\theta}|m_0^{k-1})$ is the change in free energy conditioned on just the past memories: $\rho_0(m_{k,0}) \rightarrow \rho_\theta(m_{k,\theta}|m_0^{k-1})$. Summing over all measurements, we find

$$W - \Delta\mathcal{F}(m_1^N|m_0) \geq TI_{v \rightarrow y}^N, \quad (77)$$

where $W = \sum_{k=1}^N W_k$ is the work to perform all N measurements, $\Delta\mathcal{F}(m_1^N|m_0) = \sum_{k=1}^N \Delta\mathcal{F}(m_{k,\theta}|m_0^{k-1})$ is the change in entire tape's free energy, and we have identified the discrete version of the transfer entropy [10],

$$I_{v \rightarrow y}^N = \sum_{k=1}^N I(m_{k,\theta}; v_{k-1}|m_0^{k-1}) \quad (78)$$

which is reviewed in Appendix B. Importantly, by construction, the statistics of each memory reproduce the statistics of the measurement outcomes, so equivalently

$$I_{v \rightarrow y}^N = \sum_{k=1}^N I(y_k; v_{k-1}|y_0^{k-1}) = \left\langle \ln \frac{\hat{\mathcal{P}}[y_0^N|v_0^N, y_0]}{\mathcal{P}[y_0^N|y_0]} \right\rangle. \quad (79)$$

Taking the limit as the number of measurements go to infinity while the time between them goes to zero, we obtain

$$\dot{W} - \dot{\mathcal{F}} = \lim_{t \rightarrow \infty} \frac{1}{t} \left[\lim_{\Delta s \rightarrow 0} W - \Delta\mathcal{F}(m_1^N|m_0) \right] \geq T \dot{I}_{v \rightarrow y}. \quad (80)$$

Thus, the transfer entropy rate is the minimum rate at which free energy is consumed to write to the memories. The slow protocol that we described previously saturates this bound, since it is quasistatic and therefore thermodynamically reversible.

At this point, it is worthwhile to make a connection to a class of Maxwell-demon models that exploit a tape of low entropy, auxiliary systems or cells, similar to what we have just described [24, 56, 57, 58, 59, 60]. Apart from the study in [60], these models use an ideal tape that has no internal energy, and therefore cannot exchange energy with the system, but only entropy; a setup sometimes referred to as an information reservoir [24, 58]. Under these conditions, a second-law-like inequality has been predicted that shows that the extracted work is bounded by the increase in entropy of each individual auxiliary system, ignoring the correlations between the different cells. Our memories, on the other hand, have internal energy and therefore the natural thermodynamic quantity to consider is the free energy instead of the entropy. Therefore to fit our measurement model into this tape-model framework, we must relate our information bound on the work to measure to a bound that ignores the correlations. To this end, we start with the bound for the energy to measure $W - \Delta\mathcal{F}(m_1^N|m_0) \geq T I_{v \rightarrow y}^N$ in (77), which includes through $\Delta\mathcal{F}$ all the correlations between different memories. By noting that ignoring correlations and conditioning increases the entropy, $H(m_1^N|m_0) \leq \sum_k H(m_{k,\theta})$ [42], we can conclude that ignoring the correlations decreases the free energy $\mathcal{F}(m_1^N|m_0) \geq \sum_k \mathcal{F}(m_{k,\theta})$. As a result, we have from (77) and the initial independence of each memory the series of inequalities

$$W - \sum_{k=1}^N \Delta\mathcal{F}(m_{k,\theta}) \geq T I_{v \rightarrow y}^N \geq W_{\text{ext}}. \quad (81)$$

For the ideal tape with no internal energy this reduces to $\sum_k \Delta H(m_{k,\theta}) \geq W_{\text{ext}}$ recovering the ideal-tape bound [24, 56, 57, 58, 59, 60] in our setup. Equation (81) may lead us to conclude that the bound on the extracted work from the tape-model

framework, $W - \sum_{k=1}^N \Delta \mathcal{F}(m_{k,\theta})$, is weaker than that provided by the transfer entropy. However, this would be too hasty, because these tape models allow a more general interaction between the tape cells and the system. Whereas, in our setup the memory evolution is assumed to occur separately with the velocity fixed, the tape models consider a dynamics where the memory (or cell) would be allowed to evolve simultaneously with the velocity. From this point of view, the measurement model we have presented is a special case of these more general tape models, and it is exactly our assumption that the velocity is frozen during measurement that allows us to tighten the tape-model bound using the transfer entropy. Further comparisons of such tape models with other information measures and more traditional statements of the second law can be found in [23, 24].

Finally, it should be noted that the preceding second law analysis can be viewed as a specific implementation of the information flow framework (outlined in sections 3.2 and 4.1) applied to a nonautonomously driven auxiliary memory composed of a sequence of many subsystems, see [16].

4.3. Discussion

The transfer entropy rate and information flow both bound the energy consumed during measurement. However, each measurement scenario is distinct, and in general each of these information measures will not bound the energy consumption for the other's measurement scenario. An example where $\dot{I}_{v \rightarrow y} > \dot{Q}^y/T$ is possible is presented in [14]; thus, the transfer entropy rate does not generally lower bound the amount of heat dissipated by a single memory being constantly rewritten. Our model corroborates this observation, as verified in figure 4 by the crossing of \dot{Q}^y/T and $\dot{I}_{v \rightarrow y}$. The one exception is if the controller implements the Kalman-Bucy filter (57). In which case, the equality of the information measures, $\dot{I}_{\text{flow}} = \dot{I}_{v \rightarrow y}$, implies that there is a unique lower bound to the energy required for measurement.

To conclude this section, we take a broader perspective. Our observation that the transfer entropy rate and information flow both represent the minimum (free) energy consumed (or alternatively the entropy produced) in the auxiliary memory to create that information, suggests that it is reasonable to interpret some second-law-like inequalities as actually telling us something about the thermodynamics of the system *and* its surroundings, where the surroundings include the measurement device. This allows us to incorporate information into the standard statement of the second law of thermodynamics through a kind of information reservoir on equal footing with other traditional thermodynamic reservoirs, similar to what was suggested for tapes in [24, 58]:

$$\dot{S}_i = \dot{S} + \dot{S}_{\text{env}} = \dot{S} - \frac{\dot{Q}}{T} + \dot{I} \geq 0, \quad (82)$$

which is equivalent to (1) in the steady state. Here, \dot{I} represents the minimum entropy produced in the environment that allows for feedback, with the minimum attained for reversible measurement. The appropriate choice of \dot{I} – transfer entropy rate or

information flow – depends on which type of information reservoir we wish to use. From this point of view, the efficiency ε introduced in (55) is a true measure of energetic efficiency that quantifies how faithfully the energy supplied by a reversible memory is extracted back out as work.

5. Summary

We have explored a collection of information measures that appear in second-law-like inequalities for measurement and feedback, using the tools of stochastic thermodynamics and optimal control theory. We have seen that these measures form a hierarchy of bounds on the extracted work, and that the Kalman-Bucy filter optimally will extract information and energy. Even though each measure offers a different numerical bound on the extracted work, they also each correspond to different ways of gathering information. With this distinction in mind, these second-law-like inequalities can be seen as manifestations of the second law of thermodynamics, since they include the entropy production of the system and surroundings, including the controller.

Acknowledgments

We would like to thank Martin Rosinberg for a carefully reading of this manuscript. J.M.H. is supported financially by the ARO MURI grant W911NF-11-1-0268 and H. S. is supported financially by the Swedish Research Council under grant 2013-5523.

Appendix A. Steady state probability density

The Gaussian steady state probability density in (18) is completely characterized by its means, which are zero, and the covariance matrix Σ . The elements of Σ can be determined by exploiting the Fokker-Planck equation (14) to develop a collection of equations for the variances $\langle v^2 \rangle$, $\langle y^2 \rangle$, and $\langle vy \rangle$, as described in [61] for example. The time-independent steady state solutions can then be shown to satisfy the algebraic equations

$$\begin{aligned} \sigma_y^2 - \sigma_{vy} &= \frac{\sigma^2}{2\tau} \\ \frac{\gamma}{m} \sigma_v^2 + \frac{a}{m} \sigma_{vy} &= \frac{\gamma T}{m^2} \\ \left(\frac{\gamma}{m} + \frac{1}{\tau} \right) \sigma_{vy} + \frac{a}{m} \sigma_y^2 - \frac{1}{\tau} \sigma_v^2 &= 0, \end{aligned} \tag{A.1}$$

whose solutions can be obtained after some lengthy algebra,

$$\Sigma = \begin{pmatrix} \frac{T}{m} \frac{1+(a/\gamma)(a\sigma^2/(2T))+(1+a/\gamma)(\gamma\tau/m)}{1+a/\gamma+(1+a/\gamma)(\gamma\tau/m)} & \frac{T}{m} \frac{1-a\sigma^2/(2T)}{(1+a/\gamma)(1+\gamma\tau/m)} \\ \frac{T}{m} \frac{1-a\sigma^2/(2T)}{(1+a/\gamma)(1+\gamma\tau/m)} & \frac{\sigma^2}{2\tau} \frac{1+a/\gamma+\gamma\tau/m+(2T/m)/(\sigma^2/\tau)}{(1+a/\gamma)(1+\gamma\tau/m)} \end{pmatrix}. \tag{A.2}$$

Appendix B. Path probabilities and the transfer entropy rate

In this appendix, we demonstrate how we arrive at (26) for the trajectory probability density \mathcal{P} , and how this structure allows the compact expression for the transfer entropy rate in (27).

The analysis precedes by discretizing the evolution over the time interval $s = 0$ to t into steps of width $\Delta s = t/N$ as $s_k = k\Delta t$ for $k = 0, \dots, N$ with $v_k \equiv v_{s_k}$ and $y_k \equiv y_{s_k}$. We are interested in determining the probability density $\mathcal{P}[v_0^N, y_0^N]$ to observe the pair of discrete trajectories $v_0^N = \{v_k\}_{k=0}^N$ and $y_0^N = \{y_k\}_{k=0}^N$. To this end, we discretize the Langevin equation (13) as

$$\begin{aligned} mv_{k+1} &= mv_k - (\gamma v_k + ay_k)\Delta s + \Delta\xi_k \\ \tau y_{k+1} &= \tau y_k - (y_k - v_k)\Delta s + \Delta\eta_k \end{aligned} \quad (\text{B.1})$$

where $\Delta\xi_k$ ($\Delta\eta_k$) are independent, zero-mean, Gaussian random variable with covariance $\langle \Delta\xi_k \Delta\xi_l \rangle = 2\gamma T \Delta s \delta_{kl}$ ($\langle \Delta\eta_k \Delta\eta_l \rangle = \sigma^2 \Delta s \delta_{kl}$). From this we deduce that to lowest order in Δs the transition probability splits into separate v and y evolutions as [15]

$$\begin{aligned} P(v_{k+1}, y_{k+1} | v_k, y_k) &= P(v_{k+1} | v_k, y_k) P(y_{k+1} | v_k, y_k) \\ &= \sqrt{\frac{m^2}{4\pi\gamma T \Delta s}} \exp \left[-\frac{(mv_{k+1} - mv_k + (\gamma v_k + ay_k)\Delta s)^2}{4\gamma T \Delta s} \right] \\ &\quad \times \sqrt{\frac{\tau^2}{2\pi\sigma^2 \Delta s}} \exp \left[-\frac{(\tau y_{k+1} - \tau y_k - (y_k - v_k)\Delta s)^2}{2\sigma^2 \Delta s} \right]. \end{aligned} \quad (\text{B.2})$$

Thus, the joint trajectory probability takes the form

$$\mathcal{P}[v_0^N, y_0^N] = P(v_N | v_{N-1}, y_{N-1}) P(y_N | v_{N-1}, y_{N-1}) \cdots P(v_1 | v_0, y_0) P(y_1 | v_0, y_0) p(v_0, y_0), \quad (\text{B.3})$$

with arbitrary initial density $p(v_0, y_0)$. Since the evolution naturally divides, it suggests introducing the trajectory conditional probabilities

$$\hat{\mathcal{P}}[v_0^N | y_0^N, v_0] = P(v_N | v_{N-1}, y_{N-1}) \cdots P(v_2 | v_1, y_1) P(v_1 | v_0, y_0) \quad (\text{B.4})$$

$$\hat{\mathcal{P}}[y_0^N | v_0^N, y_0] = P(y_N | v_{N-1}, y_{N-1}) \cdots P(y_2 | v_1, y_1) P(y_1 | v_0, y_0), \quad (\text{B.5})$$

in terms of which the joint trajectory probability becomes

$$\mathcal{P}[v_0^N, y_0^N] = \hat{\mathcal{P}}[v_0^N | y_0^N, v_0] \hat{\mathcal{P}}[y_0^N | v_0^N, y_0] p(v_0, y_0). \quad (\text{B.6})$$

Equations (24), (25), and (26) are the continuous time versions of the preceding equations obtained in the limit $\Delta s \rightarrow 0$.

In this discretized setup, we can directly apply the theory of discrete feedback [3, 5, 7, 14, 25, 39]. Here, the transfer entropy after N measurements is given as

$$I_{v \rightarrow y}^N = \sum_{k=0}^{N-1} \int P(v_k, y_0^{k+1}) \ln \left[\frac{P(y_{k+1} | v_k, y_k)}{P(y_{k+1} | y_0^k)} \right] dv_k dy_0^{k+1}. \quad (\text{B.7})$$

We see that the transfer entropy is the relative entropy between the transition probability of y given v , $P(y_{k+1} | v_k, y_k)$, and the unconditioned transition probability, $P(y_{k+1} | y_0^k)$, averaged over (v_k, y_0^k) . Recall that the relative entropy between two probability densities

$f(x)$ and $g(x)$ is $D(f||g) = \int f(x) \ln[f(x)/g(x)]dx$ [42]. In this way, the transfer entropy measures the affect the velocity has on the measurement dynamics, that is, how distinguishable the measurement dynamics given the velocity are from the measurement dynamics without the velocity. Expanding the sum we can rewrite (B.7) as

$$I_{v \rightarrow y}^N = \int dv_0^N dy_0^N \mathcal{P}[v_0^N, y_0^N] \ln \frac{\hat{\mathcal{P}}[y_0^N | v_0^N, y_0]}{\mathcal{P}[y_0^N | y_0]}. \quad (\text{B.8})$$

The continuous time version appears in (27).

Appendix C. Power spectra formulae for information rates

In this appendix, we sketch how entropy rates for stationary Gaussian processes can be expressed in terms of the processes' correlation functions, following the developments in [47, 48].

Let us consider a discretization with spacing $\Delta s = t/N$ of a Gaussian stochastic process $\vec{x} = \{x_k\}_{k=0}^N$. It is completely characterized by its mean $\vec{\mu} = \{\mu_k\} = \{\langle x_k \rangle\}$ and covariance matrix \mathbf{C} with elements $C_{mn} = \langle (x_m - \mu_m)(x_n - \mu_n) \rangle$, which we assume to be time-independent, $C_{mn} = c(|m - n|)$, an example being a stationary process:

$$\mathcal{P}(\vec{x}) = \frac{1}{\sqrt{(2\pi)^N |\mathbf{C}|}} \exp \left[-\frac{1}{2} (\vec{x} - \vec{\mu}) \cdot \mathbf{C}^{-1} \cdot (\vec{x} - \vec{\mu}) \right]. \quad (\text{C.1})$$

The power spectra formulae for the information rates follow from the observation that the entropy of such a Gaussian distribution is completely characterized by the covariance matrix:

$$H(\vec{x}) = \frac{N}{2} \ln(2\pi e) + \frac{1}{2} \ln |\mathbf{C}|. \quad (\text{C.2})$$

Since the process is causal, the covariance matrix has a Toeplitz structure, $C_{mn} = c(|m - n|)$, which allows us to diagonalize it in the limit $N \rightarrow \infty$ using its Fourier transform $C(\omega) = \sum_{s=0}^N e^{-i\omega s} c(s)$, with $\omega = 2\pi/t$. In which case, the entropy rate can be expressed as [48]

$$\dot{H} = \lim_{N \rightarrow \infty} \frac{1}{\Delta s N} H(\vec{x}) = \frac{1}{2\Delta s} \ln(2\pi e) + \frac{1}{2} \int_{-\pi/\Delta s}^{\pi/\Delta s} \ln C(\omega) d\omega. \quad (\text{C.3})$$

The transfer entropy is the difference in entropy rate between the trajectory of measurement outcomes $\mathcal{P}[y_0^N | y_0]$ and the entropy rate for $\hat{\mathcal{P}}[y_0^N | v_0^N, y_0]$:

$$\dot{I}_{v \rightarrow y} = \lim_{N \rightarrow \infty} \frac{1}{\Delta s N} \left[H(y_0^N | y_0) - \hat{H}(y_0^N | v_0^N, y_0) \right] = -\frac{1}{4\pi} \int_{-\pi/\Delta s}^{\pi/\Delta s} \ln \frac{\hat{C}_{yy|v}(\omega)}{C_{yy}(\omega)} d\omega. \quad (\text{C.4})$$

Taking the continuous time limit $\Delta s \rightarrow 0$, we recover the expression in (27). Similarly, the trajectory mutual information is

$$\begin{aligned} \dot{I}_{\text{traj}} &= \lim_{N \rightarrow \infty} \frac{1}{\Delta s N} [H(y_0^N) + H(v_0^N) - H(v_0^N, y_0^N)] \\ &= -\frac{1}{4\pi} \int_{-\pi/\Delta s}^{\pi/\Delta s} \ln \frac{\mathcal{C}(\omega)}{C_{vv}(\omega)C_{yy}(\omega)} d\omega, \end{aligned} \quad (\text{C.5})$$

where $\mathcal{C}(\omega)$ is the Fourier transform of the covariance matrix of the joint measurement and velocity process. One can show, as in [47], that

$$\mathcal{C}(\omega) = C_{vv}(\omega)C_{yy}(\omega) - |C_{vy}(\omega)|^2, \quad (\text{C.6})$$

which when substituted into (C.5) recovers (47) after the taking $\Delta s \rightarrow 0$.

Appendix D. Calculation of information rates

In this appendix we calculate $\dot{I}_{v \rightarrow y}$ in (28) and \dot{I}_{traj} in (47). As a first step, we must determine the Fourier transforms of various correlation functions. To this end, we begin by Fourier transforming the equations of motion for v_t and y_t in (13):

$$\begin{aligned} im\omega\hat{v}_\omega &= -\gamma\hat{v}_\omega - a\hat{y}_\omega + \hat{\xi}_\omega \\ i\tau\omega\hat{y}_\omega &= -(\hat{y}_\omega - \hat{v}_\omega - \hat{\eta}_\omega) \end{aligned} \quad (\text{D.1})$$

with $\langle |\hat{\xi}_\omega|^2 \rangle = 2\gamma T$ and $\langle |\hat{\eta}_\omega|^2 \rangle = \sigma^2$.

Let us start by determining $\dot{I}_{v \rightarrow y}$, which requires two correlation functions obtained from the solutions of (D.1) as

$$\hat{C}_{yy|v}(\omega) = \left\langle \left| \hat{y}_\omega - \frac{\hat{v}_\omega}{i\tau\omega + 1} \right|^2 \right\rangle = \frac{\sigma^2}{\tau^2\omega^2 + 1}, \quad (\text{D.2})$$

and

$$\begin{aligned} C_{yy}(\omega) = \langle |\hat{y}_\omega|^2 \rangle &= \left| \frac{1}{1 + \frac{a}{(i\tau\omega + 1)(im\omega + \gamma)}} \right|^2 \frac{1}{\tau^2\omega^2 + 1} \left(\frac{2\gamma T}{m^2\omega^2 + \gamma^2} + \sigma^2 \right) \\ &\equiv |\mathcal{S}|^2 \frac{1}{\tau^2\omega^2 + 1} \left(\frac{2\gamma T}{m^2\omega^2 + \gamma^2} + \sigma^2 \right), \end{aligned} \quad (\text{D.3})$$

where \mathcal{S} is known as the sensitivity function of the feedback system [38]. Thus, the transfer entropy rate is

$$\dot{I}_{v \rightarrow y} = -\frac{1}{4\pi} \int_{-\infty}^{\infty} \underbrace{\ln \left(\frac{\sigma^2(m^2\omega^2 + \gamma^2)}{\sigma^2(m^2\omega^2 + \gamma^2) + 2\gamma T} \right)}_A + \underbrace{\ln |\mathcal{S}|^2}_{B} d\omega. \quad (\text{D.4})$$

These integrals can be performed by exploiting the formula [62]

$$\int_0^\infty \ln \left(\frac{z^2 + a^2}{z^2 + b^2} \right) dz = \pi(a - b). \quad (\text{D.5})$$

In particular,

$$A = -\frac{1}{2\pi} \int_0^\infty \ln \left(\frac{\omega^2 + (\gamma/m)^2}{\omega^2 + (\gamma/m)^2 + 2\gamma T/(m^2\sigma^2)} \right) d\omega = \frac{\gamma}{2m} \left(\sqrt{1 + \frac{2T/\gamma}{\sigma^2}} - 1 \right), \quad (\text{D.6})$$

and

$$B = -\frac{1}{4\pi} \int_{-\infty}^{\infty} \ln |\mathcal{S}|^2 d\omega = 0, \quad (\text{D.7})$$

which recovers (29). That logarithmic integrals of the sensitivity function, such as B , equals zero holds with great generality. In fact, it represents a well-known conservation principle in control theory known as Bode's integral formula [38].

To determine \dot{I}_{traj} , we first note that $\dot{I}_{\text{traj}} = \dot{I}_{v \rightarrow y} + \dot{I}_{y \rightarrow v}$. Since we already know $\dot{I}_{v \rightarrow y}$, it remains to determine

$$\dot{I}_{y \rightarrow v} = -\frac{1}{4\pi} \int_{-\infty}^{\infty} \ln \frac{\hat{C}_{vv|y}(\omega)}{C_{vv}(\omega)} d\omega. \quad (\text{D.8})$$

The power spectra are obtained from (D.1) as

$$\begin{aligned} \hat{C}_{vv|y}(\omega) &= \left\langle \left| \hat{v}_\omega + \frac{a\hat{y}_\omega}{im\omega + \gamma} \right|^2 \right\rangle = \frac{2\gamma T}{m^2\omega^2 + \gamma^2} \\ C_{vv}(\omega) &= \langle |\hat{v}_\omega|^2 \rangle = |\mathcal{S}|^2 \frac{1}{m^2\omega^2 + \gamma^2} \left(\frac{a^2\sigma^2}{\tau^2\omega^2 + 1} + 2\gamma T \right). \end{aligned} \quad (\text{D.9})$$

Therefore, recognizing that the contribution from the sensitivity function \mathcal{S} is zero, we have

$$\dot{I}_{y \rightarrow v} = -\frac{1}{4\pi} \int_{-\infty}^{\infty} \ln \left(\frac{\omega^2 + 1/\tau^2}{\omega^2 + 1/\tau^2 + a^2\sigma^2/(2\gamma T\tau^2)} \right) = \frac{1}{2\tau} \left(\sqrt{1 + \frac{a^2\sigma^2}{2\gamma T}} - 1 \right), \quad (\text{D.10})$$

by virtue of (D.5).

References

- [1] Callen H B 1985 *Thermodynamics and an introduction to thermostatistics*, 2nd edition (John Wiley and Sons, New York)
- [2] Leff H S and Rex A F (eds) 1990 *Maxwell's Demon: Entropy, Information, Computing* (Princeton University Press, New Jersey)
- [3] Sagawa T and Ueda M 2008 *Phys. Rev. Lett.* **100** 080403
- [4] Cao F J and Feito M 2009 *Phys. Rev. E* **79** 041118
- [5] Horowitz J M and Vaikuntanathan S 2010 *Phys. Rev. E* **82** 061120
- [6] Fujitani Y and Suzuki H 2010 *J. Phys. Soc. Jap.* **79** 104003
- [7] Ponmurugan M 2010 *Phys. Rev. E* **82** 031129
- [8] Sagawa T and Ueda M 2012 *Phys. Rev. E* **85** 021104
- [9] Abreu D and Seifert U 2012 *Phys. Rev. Lett.* **108** 030601
- [10] Ito S and Sagawa T 2013 *Phys. Rev. Lett.* **111** 180603
- [11] Sandberg H, Delvenne J C, Newton N J and Mitter S K 2014 Maximum work extraction and implementation costs for non-equilibrium maxwell's demons arXiv:1402.1010
- [12] Schreiber T 2000 *Phys. Rev. Lett.* **85** 461
- [13] Barato A, Hartich D and Seifert U 2013 *Phys. Rev. E* **87** 042104
- [14] Hartich D, Barato A C and Seifert U 2014 *J. Stat. Mech.* P02016
- [15] Allahverdyan A E, Janzing D and Mahler G 2009 *J. Stat. Mech.: Theor. Exp.* P09011
- [16] Horowitz J M and Esposito M 2014 *Phys. Rev. X* **4** 031015 arXiv:1402.3276
- [17] Sagawa T and Ueda M 2012 *Phys. Rev. Lett.* **109** 180602
- [18] Sagawa T and Ueda M 2013 *New J. Phys.* **15** 125012
- [19] Kim K H and Qian H 2007 *Phys. Rev. E* **75** 022102
- [20] Munakata T and Rosinberg M L 2012 *J. Stat. Mech.* P05010
- [21] Munakata T and Rosinberg M L 2013 *J. Stat. Mech.* P06014
- [22] Munakata T and Rosinberg M L 2014 *Phys. Rev. Lett.* **112** 180601

- [23] Horowitz J M, Sagawa T and Parrondo J M R 2013 *Phys. Rev. Lett.* **111** 010602
- [24] Barato A C and Seifert U 2014 *Phys. Rev. Lett.* **112** 090601
- [25] Suzuki H and Fujitani Y 2009 *J. Phys. Soc. Jap.* **78** 074007
- [26] Abreu D and Seifert U 2011 *Europhys. Lett.* **94** 10001
- [27] Horowitz J M and Parrondo J M R 2011 *Europhys. Lett.* **95** 10005
- [28] Bauer M, Abreu D and Seifert U 2012 *J. Phys. A: Math. Theor.* **45** 162001
- [29] Horowitz J M and Parrondo J M R 2013 *Acta. Phys. Pol. B* **44** 803–814
- [30] Garnier N and Ciliberto S 2005 *Phys. Rev. E* **71** 060101(R)
- [31] Joubaud S, Garnier N B and Ciliberto S 2008 *Europhys. Lett.* **82** 30007
- [32] Kubo R, Toda M and Hashitsume N 1985 *Statistical Physics II: Nonequilibrium Statistical Mechanics* (Springer-Verlag, Berlin)
- [33] Risken H 1984 *The Fokker-Planck Equation: Methods of Solution and Applications* (Springer-Verlag, New York)
- [34] Spinney R E and Ford I J 2012 *Phys. Rev. E* **85** 051113
- [35] Tomé T and de Oliveira M 2010 *Phys. Rev. E* **82** 021120
- [36] Sekimoto K 2010 *Stochastic Energetics (Lect. Notes Phys.* vol 799) (Springer, Berlin Heidelberg)
- [37] Seifert U 2012 *Rep. Prog. Phys.* **75** 126001
- [38] Åström K J and Murray R M 2008 *Feedback systems: An introduction for scientists and engineers* (Princeton University Press, New Jersey)
- [39] Barato A C, Hartich D and Seifert U 2013 *J. Stat. Phys.* **153** 460–478
- [40] Sagawa T 2011 *J. Phys.: Conf. Ser.* **297** 012015
- [41] Shiraishi N and Sagawa T Fluctuation theorem for partially-masked nonequilibrium dynamics arXiv:1403.4018
- [42] Cover T M and Thomas J A 2006 *Elements of Information Theory* 2nd ed (Wiley-Interscience)
- [43] Ge H 2014 *Phys. Rev. E* **89** 022127
- [44] Diana G and Esposito M 2014 *J. Stat. Mech.: Theor. Exp.* P04010
- [45] Pinsker M S 1964 *Information and information stability of random variables and processes* (Holden-Day, San Francisco)
- [46] Fano R M 1968 *Transmission of information: A statistical theory of communications* (The M.I.T. Press)
- [47] Munakata T and Kamiyabu M 2006 *Eur. Phys. J. B* **53** 239–243
- [48] Tostevin F and Rein ten Wolde P 2010 *Phys. Rev. E* **81** 061917
- [49] Åström K J 2006 *Introduction to Stochastic Control Theory* Dover Books on Electrical Engineering Series (Dover Publications, Incorporated) ISBN 9780486445311
- [50] Wonham W 1968 *SIAM Journal on Control* **6** 312–326
- [51] Bucy R S and Joseph P D 1968 *Filtering for Stochastic Processes with Applications to Guidance* (Interscience Publishers, New York)
- [52] Mitter S K and Newton N J 2005 *Journal of Statistical Physics* **118** 145–176
- [53] Sagawa T and Ueda M 2010 *Phys. Rev. Lett.* **104** 090602
- [54] Esposito M and Van den Broeck C 2011 *Europhys. Lett.* **95** 40004
- [55] Deffner S and Lutz E 2012 *arXiv:1201.3888*
- [56] Mandal D and Jarzynski C 2012 *Proc. Nat. Ac. Sci.*
- [57] Mandal D, Quan H T and Jarzynski C 2013 *Phys. Rev. Lett.* **111** 030602
- [58] Deffner S and Jarzynski C 2013 *Phys. Rev. X* **3** 041003
- [59] Barato A C and Seifert U 2013 *Europhys. Lett.* **101** 60001
- [60] Hoppenau J and Engel A 2014 On the energetics of information exchange arXiv:1401.2270
- [61] Mazonka O and Jarzynski C 1999 Exactly solvable model illustrating far-from-equilibrium predictions arXiv:cond-mat/991212
- [62] Gradshteyn I S and Ryzhik I M 2007 *Table of Integrals, Series, and Products* (Elsevier Ltd., New York)